APPENDIX

1. Protein fold classification using patterns in protein sequences

   U: Protein sequences

   S: Sequence patterns in protein sequences.

   Q: Sequences known to belong to a particular fold (three dimensional arrangement in space of the amino acid chain). Example: sequences known to belong to a particular class/fold/superfamily in SCOP protein classification system. Other fold classification systems like CATH, CE, FSSP, VAST etc could also be used.

   Notes: While it is widely beleived that fold is a direct consequence of the sequence, successful prediction of the protein fold from the amino acid sequence remains a grand challenge in biology. Application of the method described here will provide the ability to use sequence patterns in new protein sequences to classify them into known folds.

2. Medical diagnostics using gene expression profile from DNA microarrays

   U: Patients, with and without disease

   S: Expression profile for each gene

   Q: Patients known to have a particular disease (clinical diagnosis).

3. Supermarket purchases

   U: Purchases (a purchase is the contents of the cart)

   S: Each item in the store

   Q: Time of day/ location/ age group/ mode of payment

4. Airline flights

   U: Commercial airline flights

   S: Airline

      Departure city

      Arrival city

Scheduled time of day/day of week/month/year of departure

Scheduled time of day/day of week/month/year of arrival

Crew compisition

Equipment

Weather at departure/arrival/enroute

Occupancy

Scheduled layover from previous flight

Scheduled flying time

Q: On time departure

On time arrival

Safety incidents

Customer satisfaction

5. Fast track security check at airports for frequent travellers

U: Travelling individuals

S: Flight attributes as in Application 4.

Individual attributes:

Age

Sex

Height

Weight

Name

Checked in bags

Carry on bags

Price of ticket

Class of travel

Mode of payment

Mode of purchase (travel agent/on line/airline)

Advance purchase

Accompanying passengers

Connecting to/from other flights

Q: Found violating security requirements

6. Automobile insurance

U: Automobile insurance policies

S: Auto: make, model, year, trim, price paid, color, condition, bought/leased

Geography: residence zip code, work zip code, commute distance

Driver: Individual characteristics as in application 5

Years of driving experience

Previous accidents - caused, involved(not caused)

Points in driver's license

insurance claim history/record

Q: Makes insurance claim

collision

damage

theft

personal injury

7. Matchmaking (example: dating service)

U: Match events, each constituting of a pair of individuals

S: Individual characteristics

Q: Satisfactory match

8. Matching of buyer-seller in on-line trading like eBay
   U: Trading events
   S: Individual characteristics
      Traded object characteristics
   Q: Difference between asking and traded price
      Satisfaction with trade
      Satisfaction with traded object

9. Advertisements on Web Portals
   U: web pages
   S: content keywords/indices
      Browser(domain, time of day, comes from)
   Q: likelihood of clicking on particular/class of ad
      viewing the page

10. Protein function classification U: Protein sequences
    S: Sequence patterns or other attributes in protein sequences.
    Q: Sequences known to belong to a particular functional fold.

11. Product Placement

12. Health/personal product selection
    U: patient/consumer, product/service
    S: Genotype/lefestyle of patient/consumer
    Q: Product/service

13. Identification of communication intercepts that constitute security threat

U: Communication intercepts

S: keywords, patterns, triggers

origin, destination of communication

medium (cell phone/land phone/email/mail)

time, day, date

Q: known to constitute security threat

pertain to some monitored activity (surveillance of known terrorist/drug trafficking networks)

14. Will a new patient respond well to known treatment?

Design/improve clinical trials by identifying and populating new patients into the non-overlapping set(s) that cause the most error.

U: Clinical patients

S: Symptoms

Individual characteristics (age, sex, location, lifestyle, etc)

Q: Patients known to respond to a particular treatment

15. Identifying the anonymous/misattributed authors of texts

U: Complete texts (example books, poetry)

S: patterns of words (example discovered by pattern discovery)

Q: Texts known to be by a particular author

of a particular genre/period

Example of relevance of this application in http://w3.research.ibm.com/visions/foster/foster.html